Proximity for Sums of Composite Functions☆

Patrick L. Combettes[a,*], Đinh Dũng[b], Bằng Công Vũ[c]

[a] *UPMC Université Paris 06, Laboratoire Jacques-Louis Lions – UMR 7598, 75005 Paris, France. Tel.: +33 1 4427 6319, Fax: +33 1 4427 7200, plc@math.jussieu.fr*
[b] *Vietnam National University, Information Technology Institute, Hanoi, Vietnam*
[c] *UPMC Université Paris 06, Laboratoire Jacques-Louis Lions – UMR 7598, 75005 Paris, France.*

Abstract

We propose an algorithm for computing the proximity operator of a sum of composite convex functions in Hilbert spaces and investigate its asymptotic behavior. Applications to best approximation and image recovery are described.

Keywords: Best approximation, convex optimization, duality, image recovery, proximity operator, proximal splitting algorithm

1. Introduction

Let $\mathcal{H}$ be a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\|$. The best approximation to a point $z \in \mathcal{H}$ from a nonempty closed convex set $C \subset \mathcal{H}$ is the point $P_C z \in C$ that satisfies $\|P_C z - z\| = \min_{x \in C} \|x - z\|$. The induced best approximation operator $P_C \colon \mathcal{H} \to C$, also called the projector onto C, plays a central role in several branches of applied mathematics [10]. If we designate by ι_C the indicator function of C, i.e.,

$$\iota_C \colon x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C, \end{cases} \tag{1.1}$$

then $P_C z$ is the solution to the minimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \iota_C(x) + \frac{1}{2}\|x - z\|^2. \tag{1.2}$$

☆The work of P. L. Combettes was supported the Agence Nationale de la Recherche under grant ANR-08-BLAN-0294-02. The work of Đinh Dũng and Bằng Công Vũ was supported by the Vietnam National Foundation for Science and Technology Development.

*Corresponding author

Now let $\Gamma_0(\mathcal{H})$ be the class of lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ such that $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\} \neq \varnothing$. In [13] Moreau observed that, for every function $f \in \Gamma_0(\mathcal{H})$, the proximal minimization problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \frac{1}{2}\|x - z\|^2 \tag{1.3}$$

possesses a unique solution, which he denoted by $\operatorname{prox}_f z$. The resulting proximity operator $\operatorname{prox}_f \colon \mathcal{H} \to \mathcal{H}$ therefore extends the notion of a best approximation operator for a convex set. This fruitful concept has become a central tool in mechanics, variational analysis, optimization, and signal processing, e.g., [1, 7, 16].

Though in certain simple cases closed-form expressions are available [7, 8, 14], computing $\operatorname{prox}_f z$ in numerical applications is a challenging task. The objective of this paper is to propose a splitting algorithm to compute proximity operators in the case when f can be decomposed as a sum of composite functions.

Problem 1.1 *Let $z \in \mathcal{H}$ and let $(\omega_i)_{1\leq i\leq m}$ be reals in $]0,1]$ such that $\sum_{i=1}^m \omega_i = 1$. For every $i \in \{1,\ldots,m\}$, let $(\mathcal{G}_i, \|\cdot\|_{\mathcal{G}_i})$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, let $g_i \in \Gamma_0(\mathcal{G}_i)$, and let $L_i\colon \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator. The problem is to*

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \sum_{i=1}^{m} \omega_i g_i(L_i x - r_i) + \frac{1}{2}\|x - z\|^2. \tag{1.4}$$

The underlying practical assumption we make is that the proximity operators $(\operatorname{prox}_{g_i})_{1\leq i\leq m}$ are implementable (to within some quantifiable error). We are therefore aiming at devising an algorithm that uses these operators separately. Let us note that such splitting algorithms are already available to solve Problem 1.1 under certain restrictions.

A) Suppose that $\mathcal{G}_1 = \mathcal{H}$, that $L_1 = \operatorname{Id}$, that the functions $(g_i)_{2\leq i\leq m}$ are differentiable everywhere with a Lipschitz continuous gradient, and that $r_i \equiv 0$. Then (1.4) reduces to the minimization of the sum of $f_1 = g_1 \in \Gamma_0(\mathcal{H})$ and of the smooth function $f_2 = \sum_{i=2}^m \omega_i g_i \circ L_i + \|\cdot - z\|^2/2$, and it can be solved by the forward-backward algorithm [8, 18].

B) The methods proposed in [4] address the case when, for every $i \in \{1,\ldots,m\}$, $\mathcal{G}_i = \mathcal{H}$, $L_i = \operatorname{Id}$, and $r_i = 0$.

C) The method proposed in [5] addresses the case when $m = 2$, $\mathcal{G}_1 = \mathcal{H}$, and $L_1 = \operatorname{Id}$, and $r_1 = 0$.

The restrictions imposed in A) are quite stringent since many problems involve at least two non-differentiable potentials. Let us also observe that since, in general, there is no explicit expression for $\operatorname{prox}_{g_i\circ L_i}$ in terms of $\operatorname{prox}_{g_i}$ and L_i, Problem 1.1 cannot be reduced to the setting described

in B). On the other hand, using a product space reformulation, we shall show that the setting described in C) can be exploited to solve Problem 1.1 using only approximate implementations of the operators $(\operatorname{prox}_{g_i})_{1\leq i\leq m}$. Our algorithm is introduced in Section 2, where we also establish its convergence properties. In Section 3, our results are applied to best approximation and image recovery problems.

Our notation is standard. $\mathcal{B}(\mathcal{H},\mathcal{G})$ is the space of bounded linear operators from $\mathcal{H}$ to a real Hilbert space $\mathcal{G}$. The adjoint of $L\in\mathcal{B}(\mathcal{H},\mathcal{G})$ is denoted by L^*. The conjugate of $f\in\Gamma_0(\mathcal{H})$ is the function $f^*\in\Gamma_0(\mathcal{H})$ defined by $f^*\colon u\mapsto\sup_{x\in\mathcal{H}}(\langle x\mid u\rangle-f(x))$. The projector onto a nonempty closed convex set $C\subset\mathcal{H}$ is denoted by P_C. The strong relative interior of a convex set $C\subset\mathcal{H}$ is

$$\operatorname{sri}C=\big\{x\in C\;\big|\;\operatorname{cone}(C-x)=\overline{\operatorname{span}}\,(C-x)\big\},\qquad \text{where}\quad \operatorname{cone}C=\bigcup_{\lambda>0}\big\{\lambda x\;\big|\;x\in C\big\},\tag{1.5}$$

and the relative interior of C is $\operatorname{ri}C=\big\{x\in C\;\big|\;\operatorname{cone}(C-x)=\operatorname{span}(C-x)\big\}$. We have $\operatorname{int}C\subset\operatorname{sri}C\subset\operatorname{ri}C\subset C$ and, if $\mathcal{H}$ is finite-dimensional, $\operatorname{ri}C=\operatorname{sri}C$. For background on convex analysis, see [19].

2. Main result

To solve Problem 1.1, we propose the following algorithm. Its main features are that each function g_i is activated individually by means of its proximity operator, and that the proximity operators can be evaluated simultaneously. It is important to stress that the functions $(g_i)_{1\leq i\leq m}$ and the operators $(L_i)_{1\leq i\leq m}$ are used at separate steps in the algorithm, which is thus fully decomposed. In addition, an error $a_{i,n}$ is tolerated in the evaluation of the ith proximity operator at iteration n.

Algorithm 2.1 For every $i\in\{1,\ldots,m\}$, let $(a_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\mathcal{G}_i$.

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\rho=\big(\max_{1\leq i\leq m}\|L_i\|\big)^{-2}\\
\varepsilon\in\left]0,\min\{1,\rho\}\right[\\
\text{For } i=1,\ldots,m\\
\left\lfloor\; v_{i,0}\in\mathcal{G}_i\right.
\end{array}
\right.\\
\text{For } n=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_n=z-\sum_{i=1}^m\omega_iL_i^*v_{i,n}\\
\gamma_n\in[\varepsilon,2\rho-\varepsilon]\\
\lambda_n\in[\varepsilon,1]\\
\text{For } i=1,\ldots,m\\
\left\lfloor\; v_{i,n+1}=v_{i,n}+\lambda_n\Big(\operatorname{prox}_{\gamma_n g_i^*}\big(v_{i,n}+\gamma_n(L_ix_n-r_i)\big)+a_{i,n}-v_{i,n}\Big).\right.
\end{array}
\right.
\end{array}
\tag{2.1}
$$

Note that an alternative implementation of (2.1) can be obtained via Moreau's decomposition formula in a real Hilbert space $\mathcal{G}$ [8, Lemma 2.10]

$$(\forall g \in \Gamma_0(\mathcal{G}))(\forall \gamma \in]0,+\infty[)(\forall v \in \mathcal{G}) \quad \operatorname{prox}_{\gamma g^*} v = v - \gamma \operatorname{prox}_{\gamma^{-1} g}(\gamma^{-1} v). \tag{2.2}$$

We now describe the asymptotic behavior of Algorithm 2.1.

Theorem 2.2 *Suppose that*

$$(r_i)_{1\leq i\leq m} \in \operatorname{sri}\big\{(L_i x - y_i)_{1\leq i\leq m} \;\big|\; x \in \mathcal{H}, (y_i)_{1\leq i\leq m} \in \times_{i=1}^m \operatorname{dom} g_i\big\} \tag{2.3}$$

and that

$$(\forall i \in \{1,\ldots,m\}) \quad \sum_{n\in\mathbb{N}} \|a_{i,n}\|_{\mathcal{G}_i} < +\infty. \tag{2.4}$$

Furthermore, let $(x_n)_{n\in\mathbb{N}}$, $(v_{1,n})_{n\in\mathbb{N}}$, ..., $(v_{m,n})_{n\in\mathbb{N}}$ be sequences generated by Algorithm 2.1. Then Problem 1.1 possesses a unique solution x and the following hold.

(i) *For every $i \in \{1,\ldots,m\}$, $(v_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $v_i \in \mathcal{G}_i$. Moreover, $(v_i)_{1\leq i\leq m}$ is a solution to the minimization problem*

$$\underset{v_1\in\mathcal{G}_1,\ldots,v_m\in\mathcal{G}_m}{\operatorname{minimize}} \; \frac{1}{2}\left\| z - \sum_{i=1}^m \omega_i L_i^* v_i \right\|^2 + \sum_{i=1}^m \omega_i\big(g_i^*(v_i) + \langle v_i \mid r_i\rangle\big), \tag{2.5}$$

and $x = z - \sum_{i=1}^m \omega_i L_i^* v_i$.

(ii) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to x.*

Proof. Set $f\colon \mathcal{H} \to]-\infty,+\infty] : x \mapsto \sum_{i=1}^m \omega_i g_i(L_i x - r_i)$. The assumptions imply that, for every $i \in \{1,\ldots,m\}$, the function $x \mapsto g_i(L_i x - r_i)$ is convex and lower semicontinuous. Hence, f is likewise. On the other hand, it follows from (2.3) that

$$(r_i)_{1\leq i\leq m} \in \big\{(L_i x - y_i)_{1\leq i\leq m} \;\big|\; x \in \mathcal{H}, (y_i)_{1\leq i\leq m} \in \times_{i=1}^m \operatorname{dom} g_i\big\} \tag{2.6}$$

and, therefore, that $\operatorname{dom} f \neq \varnothing$. Thus, $f \in \Gamma_0(\mathcal{H})$ and, as seen in (1.3), Problem 1.1 possesses a unique solution, namely $x = \operatorname{prox}_f z$.

Now let $\boldsymbol{\mathcal{H}}$ be the real Hilbert space obtained by endowing the Cartesian product $\mathcal{H}^m$ with the scalar product $\langle\cdot \mid \cdot\rangle_{\boldsymbol{\mathcal{H}}}\colon (\boldsymbol{x},\boldsymbol{y}) \mapsto \sum_{i=1}^m \omega_i \langle x_i \mid y_i\rangle$, where $\boldsymbol{x} = (x_i)_{1\leq i\leq m}$ and $\boldsymbol{y} = (y_i)_{1\leq i\leq m}$ denote generic elements in $\boldsymbol{\mathcal{H}}$. The associated norm is

$$\|\cdot\|_{\boldsymbol{\mathcal{H}}}\colon \boldsymbol{x} \mapsto \sqrt{\sum_{i=1}^m \omega_i \|x_i\|^2}. \tag{2.7}$$

Likewise, let $\boldsymbol{\mathcal{G}}$ denote the real Hilbert space obtained by endowing the Cartesian product $\mathcal{G}_1 \times \cdots \times \mathcal{G}_m$ with the scalar product and the associated norm respectively defined by

$$\langle \cdot \mid \cdot \rangle_{\boldsymbol{\mathcal{G}}} \colon (\boldsymbol{y}, \boldsymbol{z}) \mapsto \sum_{i=1}^{m} \omega_i \langle y_i \mid z_i \rangle_{\mathcal{G}_i} \quad \text{and} \quad \| \cdot \|_{\boldsymbol{\mathcal{G}}} \colon \boldsymbol{y} \mapsto \sqrt{\sum_{i=1}^{m} \omega_i \|y_i\|^2_{\mathcal{G}_i}}. \tag{2.8}$$

Define

$$\begin{cases} \boldsymbol{f} = \iota_{\boldsymbol{D}}, \quad \text{where} \quad \boldsymbol{D} = \big\{ (x, \ldots, x) \in \boldsymbol{\mathcal{H}} \;\big|\; x \in \mathcal{H} \big\} \\ \boldsymbol{g} \colon \boldsymbol{\mathcal{G}} \to \left]-\infty, +\infty\right] : \boldsymbol{y} \mapsto \sum_{i=1}^{m} \omega_i g_i(y_i) \\ \boldsymbol{L} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}} \colon \boldsymbol{x} \mapsto (L_i x_i)_{1 \leq i \leq m} \\ \boldsymbol{r} = (r_1, \ldots, r_m) \\ \boldsymbol{z} = (z, \ldots, z). \end{cases} \tag{2.9}$$

Then $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{H}})$, $\boldsymbol{g} \in \Gamma_0(\boldsymbol{\mathcal{G}})$, and $\boldsymbol{L} \in \mathcal{B}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{G}})$. Moreover, $\boldsymbol{D}$ is a closed vector subspace of $\boldsymbol{\mathcal{H}}$ with projector

$$\operatorname{prox}_{\boldsymbol{f}} = P_{\boldsymbol{D}} \colon \boldsymbol{x} \mapsto \left(\sum_{i=1}^{m} \omega_i x_i, \ldots, \sum_{i=1}^{m} \omega_i x_i \right) \tag{2.10}$$

and

$$\boldsymbol{L}^* \colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{v} \mapsto \big(L_i^* v_i \big)_{1 \leq i \leq m}. \tag{2.11}$$

Note that (2.8) and (2.7) yield

$$\begin{aligned} (\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad \|\boldsymbol{L}\boldsymbol{x}\|^2_{\boldsymbol{\mathcal{G}}} &= \sum_{i=1}^{m} \omega_i \|L_i x_i\|^2_{\mathcal{G}_i} \\ &\leq \sum_{i=1}^{m} \omega_i \|L_i\|^2 \|x_i\|^2 \\ &\leq \Big(\max_{1 \leq i \leq m} \|L_i\|^2 \Big) \sum_{i=1}^{m} \omega_i \|x_i\|^2 \\ &= \Big(\max_{1 \leq i \leq m} \|L_i\|^2 \Big) \|\boldsymbol{x}\|^2_{\boldsymbol{\mathcal{H}}}. \end{aligned} \tag{2.12}$$

Therefore,

$$\|\boldsymbol{L}\| \leq \max_{1 \leq i \leq m} \|L_i\|. \tag{2.13}$$

We also deduce from (2.3) that

$$\boldsymbol{r} \in \operatorname{sri}\big(\boldsymbol{L}(\operatorname{dom}\boldsymbol{f}) - \operatorname{dom}\boldsymbol{g}\big). \tag{2.14}$$

Furthermore, in view of (2.7) and (2.9), in the space $\boldsymbol{\mathcal{H}}$, (1.4) is equivalent to

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}}\;\; \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{L}\boldsymbol{x} - \boldsymbol{r}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_{\boldsymbol{\mathcal{H}}}^2. \tag{2.15}$$

Next, we derive from [5, Proposition 3.3] that the dual problem of (2.15) is to

$$\underset{\boldsymbol{v}\in\boldsymbol{\mathcal{G}}}{\text{minimize}}\;\; \widetilde{\boldsymbol{f}^*}(\boldsymbol{z} - \boldsymbol{L}^*\boldsymbol{v}) + \boldsymbol{g}^*(\boldsymbol{v}) + \langle \boldsymbol{v} \mid \boldsymbol{r}\rangle_{\boldsymbol{\mathcal{G}}}, \tag{2.16}$$

where $\widetilde{\boldsymbol{f}^*}\colon \boldsymbol{u} \mapsto \inf_{\boldsymbol{w}\in\boldsymbol{\mathcal{H}}}\big(\boldsymbol{f}^*(\boldsymbol{w}) + (1/2)\|\boldsymbol{u} - \boldsymbol{w}\|_{\boldsymbol{\mathcal{H}}}^2\big)$ is the Moreau envelope of $\boldsymbol{f}^*$. Since $\boldsymbol{f} = \iota_{\boldsymbol{D}}$, we have $\boldsymbol{f}^* = \iota_{\boldsymbol{D}^\perp}$. Hence, (2.7) and (2.10) yield

$$(\forall \boldsymbol{u} \in \boldsymbol{\mathcal{H}})\quad \widetilde{\boldsymbol{f}^*}(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u} - P_{\boldsymbol{D}^\perp}\boldsymbol{u}\|_{\boldsymbol{\mathcal{H}}}^2 = \frac{1}{2}\|P_{\boldsymbol{D}}\boldsymbol{u}\|_{\boldsymbol{\mathcal{H}}}^2 = \frac{1}{2}\left\|\sum_{i=1}^m \omega_i u_i\right\|^2. \tag{2.17}$$

On the other hand, (2.8) and (2.9) yield

$$(\forall \boldsymbol{v} \in \boldsymbol{\mathcal{G}})\quad \boldsymbol{g}^*(\boldsymbol{v}) = \sum_{i=1}^m \omega_i g_i^*(v_i) \quad\text{and}\quad \operatorname{prox}_{\boldsymbol{g}^*}\boldsymbol{v} = \big(\operatorname{prox}_{g_i^*} v_i\big)_{1\leq i\leq m}. \tag{2.18}$$

Altogether, it follows from (2.11), (2.17), (2.18), and (2.8), that

$$\text{(2.16) is equivalent to (2.5).} \tag{2.19}$$

Now define

$$(\forall n \in \mathbb{N})\quad \begin{cases} \boldsymbol{x}_n = (x_n, \ldots, x_n)\\ \boldsymbol{v}_n = (v_{1,n}, \ldots, v_{m,n})\\ \boldsymbol{a}_n = (a_{1,n}, \ldots, a_{m,n}). \end{cases} \tag{2.20}$$

Then, in view of (2.9), (2.10), (2.11), (2.13), and (2.18), (2.1) is a special case of the following routine.

$$\begin{array}{l}
\text{Initialization}\\
\left\lfloor\begin{array}{l}
\rho = \|\boldsymbol{L}\|^{-2}\\
\varepsilon \in \left]0, \min\{1,\rho\}\right[\\
\boldsymbol{v}_0 \in \boldsymbol{\mathcal{G}}
\end{array}\right.\\
\text{For } n = 0, 1, \ldots\\
\left\lfloor\begin{array}{l}
\boldsymbol{x}_n = \operatorname{prox}_{\boldsymbol{f}}(\boldsymbol{z} - \boldsymbol{L}^*\boldsymbol{v}_n)\\
\gamma_n \in [\varepsilon, 2\rho - \varepsilon]\\
\lambda_n \in [\varepsilon, 1]\\
\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + \lambda_n\big(\operatorname{prox}_{\gamma_n\boldsymbol{g}^*}(\boldsymbol{v}_n + \gamma_n(\boldsymbol{L}\boldsymbol{x}_n - \boldsymbol{r})) + \boldsymbol{a}_n - \boldsymbol{v}_n\big).
\end{array}\right.
\end{array} \tag{2.21}$$

Moreover, (2.4) implies that $\sum_{n\in\mathbb{N}} \|\boldsymbol{a}_n\|_{\boldsymbol{\mathcal{G}}} < +\infty$. Hence, it follows from (2.14) and [5, Theorem 3.7] that the following hold, where $\boldsymbol{x}$ is the solution to (2.15).

(a) $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ converges weakly to a solution $\boldsymbol{v}$ to (2.16) and $\boldsymbol{x} = \operatorname{prox}_{\boldsymbol{f}}(\boldsymbol{z} - \boldsymbol{L}^*\boldsymbol{v})$.

(b) $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges strongly to $\boldsymbol{x}$.

In view of (2.7), (2.8), (2.9), (2.10), (2.11), (2.19), and (2.20), items (a) and (b) provide respectively items (i) and (ii). □

Remark 2.3 Let us consider Problem 1.1 in the special case when $(\forall i \in \{1,\ldots,m\})$ $\mathcal{G}_i = \mathcal{H}$, $L_i = \operatorname{Id}$, and $r_i = 0$. Then (1.4) reduces to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \sum_{i=1}^{m} \omega_i g_i(x) + \frac{1}{2}\|x - z\|^2. \tag{2.22}$$

Now let us implement Algorithm 2.1 with $\gamma_n \equiv 1$, $\lambda_n \equiv 1$, $a_{i,n} \equiv 0$, and $v_{i,0} \equiv 0$. The iteration process resulting from (2.1) can be written as

$$\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
x_0 = z\\
\text{For } i = 1,\ldots,m\\
\left\lfloor\; v_{i,0} = 0\right.
\end{array}\right.\\
\text{For } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\text{For } i = 1,\ldots,m\\
\left\lfloor\; v_{i,n+1} = \operatorname{prox}_{g_i^*}(x_n + v_{i,n}).\right.\\
x_{n+1} = z - \sum_{i=1}^{m} \omega_i v_{i,n+1}.
\end{array}\right.
\end{array} \tag{2.23}$$

For every $i \in \{1,\ldots,m\}$ and $n \in \mathbb{N}$, set $z_{i,n} = x_n + v_{i,n}$. Then (2.23) yields

$$\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
x_0 = z\\
\text{For } i = 1,\ldots,m\\
\left\lfloor\; z_{i,0} = z\right.
\end{array}\right.\\
\text{For } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_{n+1} = z - \sum_{i=1}^{m} \omega_i \operatorname{prox}_{g_i^*} z_{i,n}\\
\text{For } i = 1,\ldots,m\\
\left\lfloor\; z_{i,n+1} = x_{n+1} + \operatorname{prox}_{g_i^*} z_{i,n}.\right.
\end{array}\right.
\end{array} \tag{2.24}$$

Next we observe that $(\forall n \in \mathbb{N})$ $\sum_{i=1}^{m} \omega_i z_{i,n} = z$. Indeed, the identity is clearly satisfied for $n = 0$ and, for every $n \in \mathbb{N}$, (2.24) yields $\sum_{i=1}^{m} \omega_i z_{i,n+1} = x_{n+1} + \sum_{i=1}^{m} \omega_i \operatorname{prox}_{g_i^*} z_{i,n} = (z -$

$\sum_{i=1}^m \omega_i \operatorname{prox}_{g_i^*} z_{i,n}) + \sum_{i=1}^m \omega_i \operatorname{prox}_{g_i^*} z_{i,n} = z$. Thus, invoking (2.2) with $\gamma = 1$, we can rewrite (2.24) as

$$
\begin{array}{l}
\text{Initialization} \\
\left\lfloor
\begin{array}{l}
x_0 = z \\
\text{For } i = 1, \ldots, m \\
\left\lfloor\; z_{i,0} = z \right.
\end{array}
\right. \\
\text{For } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_{n+1} = \sum_{i=1}^m \omega_i \operatorname{prox}_{g_i} z_{i,n} \\
\text{For } i = 1, \ldots, m \\
\left\lfloor\; z_{i,n+1} = x_{n+1} + z_{i,n} - \operatorname{prox}_{g_i} z_{i,n}. \right.
\end{array}
\right.
\end{array}
\tag{2.25}
$$

This is precisely the Dykstra-like algorithm proposed in [4, Theorem 4.2] for computing $\operatorname{prox}_{\sum_{i=1}^m \omega_i g_i} z$ (which itself extends the classical parallel Dykstra algorithm for projecting z onto an intersection of closed convex sets [2, 11]). Hence, Algorithm 2.1 can be viewed as an extension of this algorithm, which was derived and analyzed with different techniques in [4].

3. Applications

As noted in the Introduction, special cases of Problem 1.1 have already been considered in the literature under certain restrictions on the number m of composite functions, the complexity of the linear operators $(L_i)_{1 \leq i \leq m}$, and/or the smoothness of the potentials $(g_i)_{1 \leq i \leq m}$ (one will find specific applications in [3, 5, 7, 8, 9, 15] and the references therein). The proposed framework makes it possible to remove these restrictions simultaneously. In this section, we provide two illustrations.

3.1. Best approximation from an intersection of composite convex sets

In this section, we consider the problem of finding the best approximation $P_D z$ to a point $z \in \mathcal{H}$ from a closed convex subset D of $\mathcal{H}$ defined as an intersection of affine inverse images of closed convex sets.

Problem 3.1 *Let $z \in \mathcal{H}$ and, for every $i \in \{1, \ldots, m\}$, let $(\mathcal{G}_i, \| \cdot \|_{\mathcal{G}_i})$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, let C_i be a nonempty closed convex subset of $\mathcal{G}_i$, and let $0 \neq L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i)$. The problem is to*

$$
\underset{x \in D}{\text{minimize}}\; \|x - z\|, \quad \textit{where} \quad D = \bigcap_{i=1}^m \big\{ x \in \mathcal{H} \;\big|\; L_i x \in r_i + C_i \big\}. \tag{3.1}
$$

In view of (1.1), Problem 3.1 is a special case of Problem 1.1, where $(\forall i \in \{1, \ldots, m\})\; g_i = \iota_{C_i}$ and $\omega_i = 1/m$. It follows that, for every $i \in \{1, \ldots, m\}$ and every $\gamma \in \left]0, +\infty\right[$, $\operatorname{prox}_{\gamma g_i}$ reduces to the projector P_{C_i} onto C_i. Hence, using (2.2), we can rewrite Algorithm 2.1 in the following form, where we have set $c_{i,n} = -\gamma_n^{-1} a_{i,n}$ for simplicity.

Algorithm 3.2 For every $i \in \{1,\dots,m\}$, let $(c_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\mathcal{G}_i$.

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\rho = \big(\max_{1\le i\le m}\|L_i\|\big)^{-2}\\
\varepsilon \in \left]0,\min\{1,\rho\}\right[\\
\text{For } i=1,\dots,m\\
\left\lfloor \; v_{i,0}\in\mathcal{G}_i \right.
\end{array}
\right.\\
\text{For } n=0,1,\dots\\
\left\lfloor
\begin{array}{l}
x_n = z - \sum_{i=1}^m \omega_i L_i^* v_{i,n}\\
\gamma_n \in [\varepsilon, 2\rho-\varepsilon]\\
\lambda_n \in [\varepsilon,1]\\
\text{For } i=1,\dots,m\\
\left\lfloor \; v_{i,n+1} = v_{i,n} + \gamma_n\lambda_n\Big(L_i x_n - r_i - P_{C_i}\big(\gamma_n^{-1}v_{i,n} + L_i x_n - r_i\big) - c_{i,n}\Big). \right.
\end{array}
\right.
\end{array}
\tag{3.2}
$$

In the light of the above, we obtain the following application of Theorem 2.2(ii).

Corollary 3.3 *Suppose that*

$$
(r_i)_{1\le i\le m} \in \operatorname{sri}\big\{(L_i x - y_i)_{1\le i\le m} \;\big|\; x\in\mathcal{H}, (y_i)_{1\le i\le m} \in \times_{i=1}^m C_i\big\} \tag{3.3}
$$

and that $(\forall i \in \{1,\dots,m\})\ \sum_{n\in\mathbb{N}} \|c_{i,n}\|_{\mathcal{G}_i} < +\infty$. *Then every sequence* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm 3.2 converges strongly to the solution* $P_D z$ *to Problem 3.1.*

3.2. Nonsmooth image recovery

A wide range of signal and image recovery problems can be modeled as instances of Problem 1.1. In this section, we focus on the problem of recovering an image $\overline{x} \in \mathcal{H}$ from p noisy measurements

$$
r_i = T_i\overline{x} + s_i, \quad 1 \le i \le p. \tag{3.4}
$$

In this model, the ith measurement r_i lies in a Hilbert space $\mathcal{G}_i$, $T_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$ is the data formation operator, and $s_i \in \mathcal{G}_i$ is the realization of a noise process. A typical data fitting potential in such models is the function

$$
x \mapsto \sum_{i=1}^p \omega_i g_i(T_i x - r_i), \quad \text{where} \quad 0 \le g_i \in \Gamma_0(\mathcal{G}_i) \quad \text{and } g_i \text{ vanishes only at } 0. \tag{3.5}
$$

The proposed framework can handle $p \ge 1$ nondifferentiable functions $(g_i)_{1\le i\le p}$ as well as the incorporation of additional potential functions to model prior knowledge on the original image $\overline{x}$. In the illustration we provide below, the following is assumed.

- The image space is $\mathcal{H} = \mathrm{H}_0^1(\Omega)$, where Ω is a nonempty bounded open domain in $\mathbb{R}^2$.
- $\overline{x}$ admits a sparse decomposition in an orthonormal basis $(e_k)_{k\in\mathbb{N}}$ of $\mathcal{H}$. As discussed in [9, 20] this property can be promoted by the "elastic net" potential $x \mapsto \sum_{k\in\mathbb{N}} \phi_k(\langle x \mid e_k\rangle)$, where $(\forall k \in \mathbb{N})$ $\phi_k \colon \xi \mapsto \alpha|\xi| + \beta|\xi|^2$, with $\alpha > 0$ and $\beta > 0$. More general choices of suitable functions $(\phi_k)_{k\in\mathbb{N}}$ are available [6].
- $\overline{x}$ is piecewise smooth. This property is promoted by the total variation potential $\mathrm{tv}(x) = \int_\Omega |\nabla x(\omega)|_2 d\omega$, where $|\cdot|_2$ denotes the Euclidean norm on $\mathbb{R}^2$ [17].

Upon setting $g_i \equiv \|\cdot\|_{\mathcal{G}_i}$ in (3.5), these considerations lead us to the following formulation (see [5, Example 2.10] for more general nonsmooth potentials).

Problem 3.4 *Let $\mathcal{H} = \mathrm{H}_0^1(\Omega)$, where $\Omega \subset \mathbb{R}^2$ is nonempty, bounded, and open, let $(\omega_i)_{1\le i\le p+2}$ be reals in $]0,1]$ such that $\sum_{i=1}^{p+2} \omega_i = 1$, and let $(e_k)_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$. For every $i \in \{1,\dots,p\}$, let $0 \neq T_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$, where $(\mathcal{G}_i, \|\cdot\|_{\mathcal{G}_i})$ is a real Hilbert space, and let $r_i \in \mathcal{G}_i$. The problem is to*

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \sum_{i=1}^{p} \omega_i \|T_i x - r_i\|_{\mathcal{G}_i} + \sum_{k\in\mathbb{N}} \left(\omega_{p+1}|\langle x \mid e_k\rangle| + \frac{1}{2}|\langle x \mid e_k\rangle|^2\right) + \omega_{p+2}\,\mathrm{tv}(x). \tag{3.6}$$

It follows from Parseval's identity that Problem 3.4 is a special case of Problem 1.1 in $\mathcal{H} = \mathrm{H}_0^1(\Omega)$ with $m = p+2$, $z = 0$, and

$$\begin{cases} g_i = \|\cdot\|_{\mathcal{G}_i} \text{ and } L_i = T_i, \text{ if } 1 \le i \le p;\\ \mathcal{G}_{p+1} = \ell^2(\mathbb{N}),\ g_{p+1} = \|\cdot\|_{\ell^1},\ r_{p+1} = 0, \text{ and } L_{p+1} \colon x \mapsto (\langle x \mid e_k\rangle)_{k\in\mathbb{N}};\\ \mathcal{G}_{p+2} = \mathrm{L}^2(\Omega) \oplus \mathrm{L}^2(\Omega),\ g_{p+2} \colon y \mapsto \int_\Omega |y(\omega)|_2 d\omega,\ r_{p+2} = 0, \text{ and } L_{p+2} = \nabla. \end{cases} \tag{3.7}$$

To implement Algorithm 2.1, it suffices to note that $L_{p+1}^* \colon (\nu_k)_{k\in\mathbb{N}} \mapsto \sum_{k\in\mathbb{N}} \nu_k e_k$ and $L_{p+2}^* = -\operatorname{div}$, and to specify the proximity operators of the functions $(\gamma g_i^*)_{1\le i\le m}$, where $\gamma \in \,]0,+\infty[$. First, let $i \in \{1,\dots,p\}$. Then $g_i = \|\cdot\|_{\mathcal{G}_i}$ and therefore $g_i^* = \iota_{B_i}$, where B_i is the closed unit ball of $\mathcal{G}_i$. Hence $\mathrm{prox}_{\gamma g_i^*} = P_{B_i}$. Next, it follows from (2.2) and [8, Example 2.20] that $\mathrm{prox}_{\gamma g_{p+1}^*} \colon (\xi_k)_{k\in\mathbb{N}} \mapsto (P_{[-1,1]}\xi_k)_{k\in\mathbb{N}}$. Finally, since g_{p+2} is the support function of the set [12]

$$K = \big\{ y \in \mathcal{G}_{p+2} \;\big|\; |y|_2 \le 1 \text{ a.e.} \big\}, \tag{3.8}$$

$g_{p+2}^* = \iota_K$ and therefore $\mathrm{prox}_{\gamma g_{p+2}^*} = P_K$, which is straightforward to compute. Altogether, as $\|L_{p+1}\| = 1$ and $\|L_{p+2}\| \le 1$, Algorithm 2.1 assumes the following form (since all the proximity operators can be implemented with simple projections, we dispense with the errors terms).

Algorithm 3.5

$$
\begin{array}{l}
\text{Initialization}\\
\left\lfloor
\begin{array}{l}
\rho = \big(\max\{1, \|T_1\|, \ldots, \|T_p\|\}\big)^{-2}\\
\varepsilon \in\,]0, \min\{1, \rho\}[\\
\text{For } i = 1, \ldots, p\\
\left\lfloor\; v_{i,0} \in \mathcal{G}_i \right.\\
v_{p+1,0} = (\nu_{k,0})_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})\\
v_{p+2,0} \in \mathrm{L}^2(\Omega) \oplus \mathrm{L}^2(\Omega)
\end{array}\right.\\
\text{For } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
x_n = z - \sum_{i=1}^p \omega_i T_i^* v_{i,n} - \omega_{p+1} \sum_{k\in\mathbb{N}} \nu_{k,n} e_k + \omega_{p+2} \operatorname{div} v_{p+2,n}\\
\gamma_n \in [\varepsilon, 2\rho - \varepsilon]\\
\lambda_n \in [\varepsilon, 1]\\
\text{For } i = 1, \ldots, p\\
\left\lfloor\; v_{i,n+1} = v_{i,n} + \lambda_n \Big(\dfrac{v_{i,n} + \gamma_n (T_i x_n - r_i)}{\max\{1, \|v_{i,n} + \gamma_n (T_i x_n - r_i)\|_{\mathcal{G}_i}\}} - v_{i,n} \Big) \right.\\
\text{For every } k \in \mathbb{N},\ \nu_{k,n+1} = \nu_{k,n} + \lambda_n \Big(\dfrac{\nu_{k,n} + \gamma_n \langle x_n \mid e_k\rangle}{\max\{1, |\nu_{k,n} + \gamma_n \langle x_n \mid e_k\rangle|\}} - \nu_{k,n} \Big)\\
\text{For almost every } \omega \in \Omega,\\
\qquad v_{p+2,n+1}(\omega) = v_{p+2,n}(\omega) + \lambda_n \Big(\dfrac{v_{p+2,n}(\omega) + \gamma_n \nabla x_n(\omega)}{\max\{1, |v_{p+2,n}(\omega) + \gamma_n \nabla x_n(\omega)|_2\}} - v_{p+2,n}(\omega) \Big).
\end{array}\right.
\end{array}
\tag{3.9}
$$

Let us establish the main convergence property of this algorithm.

Corollary 3.6 *Every sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 3.5 converges strongly to the solution to Problem 3.4.*

Proof. In view of the above discussion and of Theorem 2.2(ii), it remains to check that (2.3) is satisfied. Set $S = \big\{(L_i x - y_i)_{1\le i\le m} \;\big|\; x \in \mathcal{H}, (y_i)_{1\le i\le m} \in \times_{i=1}^m \operatorname{dom} g_i\big\}$. We have $\operatorname{dom} g_i = \mathcal{G}_i$ for every $i \in \{1, \ldots, p\}$, $\operatorname{dom} g_{p+1} = \ell^1(\mathbb{N})$, and $\operatorname{dom} g_{p+2} = \mathrm{L}^2(\Omega) \oplus \mathrm{L}^2(\Omega)$. Consequently,

$$
\begin{aligned}
S &= \Big\{ (T_1 x - y_1, \ldots, T_p x - y_p, (\langle x \mid e_k\rangle - \eta_k)_{k\in\mathbb{N}}, \nabla x - y_{p+2} \;\Big|\\
&\qquad\qquad x \in \mathcal{H}, (y_i)_{1\le i\le p} \in \times_{i=1}^p \mathcal{G}_i, (\eta_k)_{k\in\mathbb{N}} \in \ell^1(\mathbb{N}), y_{p+2} \in \mathrm{L}^2(\Omega) \oplus \mathrm{L}^2(\Omega) \Big\}\\
&= \big(\times_{i=1}^p \mathcal{G}_i\big) \times \ell^2(\mathbb{N}) \times \big(\mathrm{L}^2(\Omega) \oplus \mathrm{L}^2(\Omega)\big)\\
&= \times_{i=1}^m \mathcal{G}_i.
\end{aligned}
\tag{3.10}
$$

Hence, we trivially have $(r_1, \ldots, r_p, 0, 0) \in \operatorname{sri} S$. ☐

Let us emphasize that a novelty of the above variational framework is to perform total variation image recovery in the presence of several nondifferentiable composite terms, with guaranteed strong convergence to the solution to the problem, and with elementary steps in the form of simple projections. The finite-dimensional version of the algorithm can easily be obtained by discretizing the operators ∇ and div as in [3] (see also [5, Section 4.4] for variants of the total variation potential).

References

[1] P. Alart, O. Maisonneuve, and R. T. Rockafellar (Eds.), *Nonsmooth Mechanics and Analysis – Theoretical and Numerical Advances.* Springer-Verlag, New York, 2006.

[2] H. H. Bauschke and J. M. Borwein, Dykstra's alternating projection algorithm for two sets, *J. Approx. Theory,* vol. 79, pp. 418–443, 1994.

[3] A. Chambolle, Total variation minimization and a class of binary MRF model, *Lecture Notes in Comput. Sci.,* vol. 3757, pp 136–152, 2005.

[4] P. L. Combettes, Iterative construction of the resolvent of a sum of maximal monotone operators, *J. Convex Anal.,* vol. 16, pp. 727–748, 2009.

[5] P. L. Combettes, Đinh Dũng, and B. C. Vũ, Dualization of signal recovery problems, *Set-Valued Variational Anal.,* to appear. `http://www.ann.jussieu.fr/∼plc/svva1.pdf`

[6] P. L. Combettes and J.-C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, *SIAM J. Optim.,* vol. 18, pp. 1351–1376, 2007.

[7] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* (H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, H. Wolkowicz, Eds.). Springer-Verlag, New York, 2010. `http://www.ann.jussieu.fr/∼plc/prox.pdf`

[8] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.,* vol. 4, pp. 1168–1200, 2005.

[9] C. De Mol, E. De Vito, and L. Rosasco, Elastic-net regularization in learning theory, *J. Complexity,* vol. 25, pp. 201–230, 2009.

[10] F. Deutsch, *Best Approximation in Inner Product Spaces.* Springer-Verlag, New York, 2001.

[11] N. Gaffke and R. Mathar, A cyclic projection algorithm via duality, *Metrika,* vol. 36, pp. 29–54, 1989.

[12] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d'Orsay, no. 80.01). Orsay, France, Université de Paris-XI, 1980.

[13] J.-J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 2897–2899, 1962.

[14] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France,* vol. 93, pp. 273-299, 1965.

[15] L. C. Potter and K. S. Arun, A dual approach to linear inverse problems with convex constraints, *SIAM J. Control Optim.,* vol. 31, pp. 1080–1092, 1993.

[16] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis,* 3rd printing. Springer-Verlag, New York, 2009.

[17] L. I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D,* vol. 60, pp. 259–268, 1992.

[18] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.,* vol. 29, pp. 119–138, 1991.

[19] C. Zălinescu, *Convex Analysis in General Vector Spaces,* World Scientific, River Edge, NJ, 2002.

[20] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B Stat. Methodol.,* vol. 67, pp. 301–320, 2005.